

ARTICLE

The Greeks in the West: genetic signatures of the Hellenic colonisation in southern Italy and Sicily

Sergio Tofanelli^{1,18}, Francesca Brisighelli^{2,3,18}, Paolo Anagnostou^{4,5}, George BJ Busby^{2,6}, Gianmarco Ferri⁷, Mark G Thomas⁸, Luca Taglioli¹, Igor Rudan⁹, Tatijana Zemunik¹⁰, Caroline Hayward¹¹, Deborah Bolnick¹¹, Valentino Romano^{12,13}, Francesco Cali¹², Donata Luiselli¹⁴, Gillian B Shepherd¹⁵, Sebastiano Tusa¹⁶, Antonino Facella¹⁷ and Cristian Capelli^{*,2}

Greek colonisation of South Italy and Sicily (*Magna Graecia*) was a defining event in European cultural history, although the demographic processes and genetic impacts involved have not been systematically investigated. Here, we combine high-resolution surveys of the variability at the uni-parentally inherited Y chromosome and mitochondrial DNA in selected samples of putative source and recipient populations with forward-in-time simulations of alternative demographic models to detect signatures of that impact. Using a subset of haplotypes chosen to represent historical sources, we recover a clear signature of Greek ancestry in East Sicily compatible with the settlement from Euboea during the Archaic Period (eighth to fifth century BCE). We inferred moderate sex-bias in the numbers of individuals involved in the colonisation: a few thousand breeding men and a few hundred breeding women were the estimated number of migrants. Last, we demonstrate that studies aimed at quantifying Hellenic genetic flow by the proportion of specific lineages surviving in present-day populations may be misleading. *European Journal of Human Genetics* advance online publication, 15 July 2015; doi:10.1038/ejhg.2015.124

INTRODUCTION

'...board a fast ship to Sicily, where you could sell them for a profit.'
(Homer, *Odyssey* XX 382–383)

From the eighth century BCE, the Western Mediterranean witnessed a settlement process which originated in the Aegean and had its centre of gravity in Eastern Sicily and Southern Italy, an area later known as *Magna Graecia*. A substantial agreement on the drivers of this colonisation and the geo-historical dynamics of the settlements is present^{1–3} but the nature of early settlements, the scale of demographic impact and its genetic legacy are still debated. Archaeologists, historians and demographers have proposed different degrees of Greek contribution, with scenarios ranging from a colonisation process based on small groups of males moderately admixing with autochthonous groups (Boardman⁴ p. 163; Yntema⁵), 'like ants or frogs about a pond' (Plato, *Phaedo* 109a–b), to substantial migrations from Greece and a Hellenic origin for a significant part of the pre-Roman Italian population.^{6–8} The pattern of genetic variation observed in Italian populations has been interpreted either as supporting a substantial Greek contribution to the current Sicilian and southern Italian gene pool,^{9,10} or having being shaped by other demographic processes.^{11–13}

When the genetic impact of the Greek Colonisation (GC) has been specifically addressed^{14,15} a lineage-based interpretative approach was used, relied on the frequency of the more frequent haplogroups (E-V13) or STR motifs (Balkan Modal Haplotype) in present-day Greeks. However, these approaches can be strongly misleading. Population patterns might not hold when single nucleotide polymorphisms defining finer haplogroup assignments are genotyped. Moreover, it is problematic to treat specific lineages or haplotypes as markers of GC, as these studies have done, because (1) the region where a haplotype is most common today is not necessarily the region where it originated,^{16,17} (2) modern population samples from the hypothesised source region may not be a good proxy for ancestral source populations, and (3) present-day patterns might be related to other events that triggered migration along the same route, most notably the Neolithic agricultural revolution or migratory flows during the Bronze Age, the Classical Era and the Christian Era. In addition, previous investigations have not formally tested alternative demographic models to clarify the scale of migration associated with the GC, an issue that has been puzzling demographers for long.^{18–21}

To search and characterize the genetic echoes of the demographic impact of the colonisation process, we applied a hypothesis-testing approach consisting of a deep molecular characterisation at male

¹Dipartimento di Biologia, Università di Pisa, Pisa, Italy; ²Department of Zoology, University of Oxford, Oxford, UK; ³Sezione di Medicina Legale-Istituto di Sanità Pubblica, Università Cattolica del Sacro Cuore, Roma, Italy; ⁴Dipartimento di Biologia Ambientale, Università "La Sapienza", Roma, Italy; ⁵Istituto Italiano di Antropologia, Roma, Italy; ⁶Wellcome Trust Centre for Human Genetics, Oxford, UK; ⁷Dipartimento ad Attività Integrata di Laboratori, Anatomia Patologica, Medicina Legale, U.O. Struttura Complessa di Medicina Legale, Università di Modena e Reggio Emilia, Modena, Italy; ⁸Department of Genetics, Evolution and Environment, University College London, London, UK; ⁹Centre for Population Health Sciences, The University of Edinburgh Medical School, Scotland, UK; ¹⁰Department of Medical Biology, University of Split, School of Medicine, Split, Croatia; ¹¹MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine (IGMM), University of Edinburgh, Western General Hospital, Edinburgh, UK; ¹²Dipartimento di Fisica e Chimica, Università di Palermo, Palermo, Italy; ¹³Laboratorio di Genetica Molecolare, I.R.C.C.S. Associazione Oasi Maria SS., Troina, Italy; ¹⁴Laboratorio di Antropologia Molecolare, Dipartimento di Scienze Biologiche, Geologiche e Ambientali, Università di Bologna, Bologna, Italy; ¹⁵A.D. Trendall Research Centre for Ancient Mediterranean Studies, La Trobe University, Melbourne, Victoria, Australia; ¹⁶Soprintendenza del mare, Palermo, Italy; ¹⁷Soprintendenza per i Beni Archeologici della Calabria, Reggio Calabria, Italy
*Correspondence: Professor C Cristian, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. Tel: +44 1865 271261; Fax: +44 1865 281843; E-mail: cristian.capelli@zoo.ox.ac.uk

¹⁸These authors contributed equally to this work.

Received 19 December 2014; revised 29 April 2015; accepted 6 May 2015

specific Y chromosome regions coupled with a dedicated sampling strategy to include relevant source and recipient populations. In addition, extensive simulations aimed at comparing alternative models for the origin and fate of Greek genetic contributions to southern Italy and Sicily were performed. In doing so, we: (a) tested for the presence of specific and robust signatures of Greek contribution to Italy and Sicily; (b) tested alternative models to estimate the most likely scale of the colonisation process, with a focus on the relative contribution of males *versus* females; and (c) evaluated the implications of lineage-based approaches in the characterisation of past demographic events.

MATERIALS AND METHODS

Sampling

A total of 811 unrelated individuals native to Greek districts thought to be at the origin of the GC migrations (Euboea island and Corinthia) and of cities placed in close proximity to primary and secondary Greek colonies in southern Italy were recruited through informed consent (Figure 1). The project received ethical approval by the institutions involved in the collection of the samples. Other putative recipient population samples from northern and central Italy were made available ($N=201$; Brisighelli *et al*¹²). In addition, DNA samples from Albania,²² Croatia²³ and West Anatolia²⁴ were obtained as reference populations. To analyse samples of source, recipient and reference populations

of comparable size, Italian populations were clustered into six main groups following geographical and historical criteria (Supplementary Table S1).

DNA was extracted from whole blood or oral samples (buccal swabs or saliva) through a modified salting out procedure²⁵ or commercially available extraction kits (Master Pure Purification kit, Epicentre, Nucleon BACC, GE Healthcare Bio-Sciences, Pittsburgh, PA, USA).

Genotyping

A total of 59 previously published single nucleotide polymorphisms was analysed following a hierarchical genotyping strategy. Samples were amplified in a standard PCR reaction and the SNaPshot Multiplex System (Life Technologies Corp., Carlsbad, CA, USA) primer extension protocol was used. All samples were first genotyped for markers, E-M35, F-M89, G-M201, H-M282, I-M170, K-M9, J-M172, J-M267, J-M304, R-M173, P-M45, R-M17 and R-M269, to classify them into major European branches. Samples belonging to haplogroups E-M35, E-M78, J-M172, I-M170 and R-M269 were further analysed by means of haplogroup specific multiplexes (Supplementary Table S2). Furthermore, samples assigned to haplogroup G-M201 were analysed for markers M406 and P15 through direct sequencing (Supplementary Table S3). Nomenclature used for haplogroup labelling follows YCC conventions²⁶ and recent updating (ISOGG Y-Tree 2015 <http://www.isogg.org/tree/>).

The entire data set was also analysed at a total of 26 Y chromosome short tandem repeats (YSTRs): those included in the AmpFISTR Yfiler PCR

1. Akragas
2. Dikaiarchia
3. Elea
4. Gela
5. Himera
6. Hipponion
7. Kamarina
8. Katane
9. Kaulonia
10. Kroton
11. Kyme
12. Laos
13. Leontinoi
14. Lipara
15. Lokroi
16. Medma
17. Megara Hyblaea
18. Metapontion
19. Naxos
20. Partenope
21. Pithekoussai
22. Poseidonia
23. Rhegion
24. Selinuntis
25. Siris
26. Sybaris
27. Syracuse
28. Taras
29. Zancle

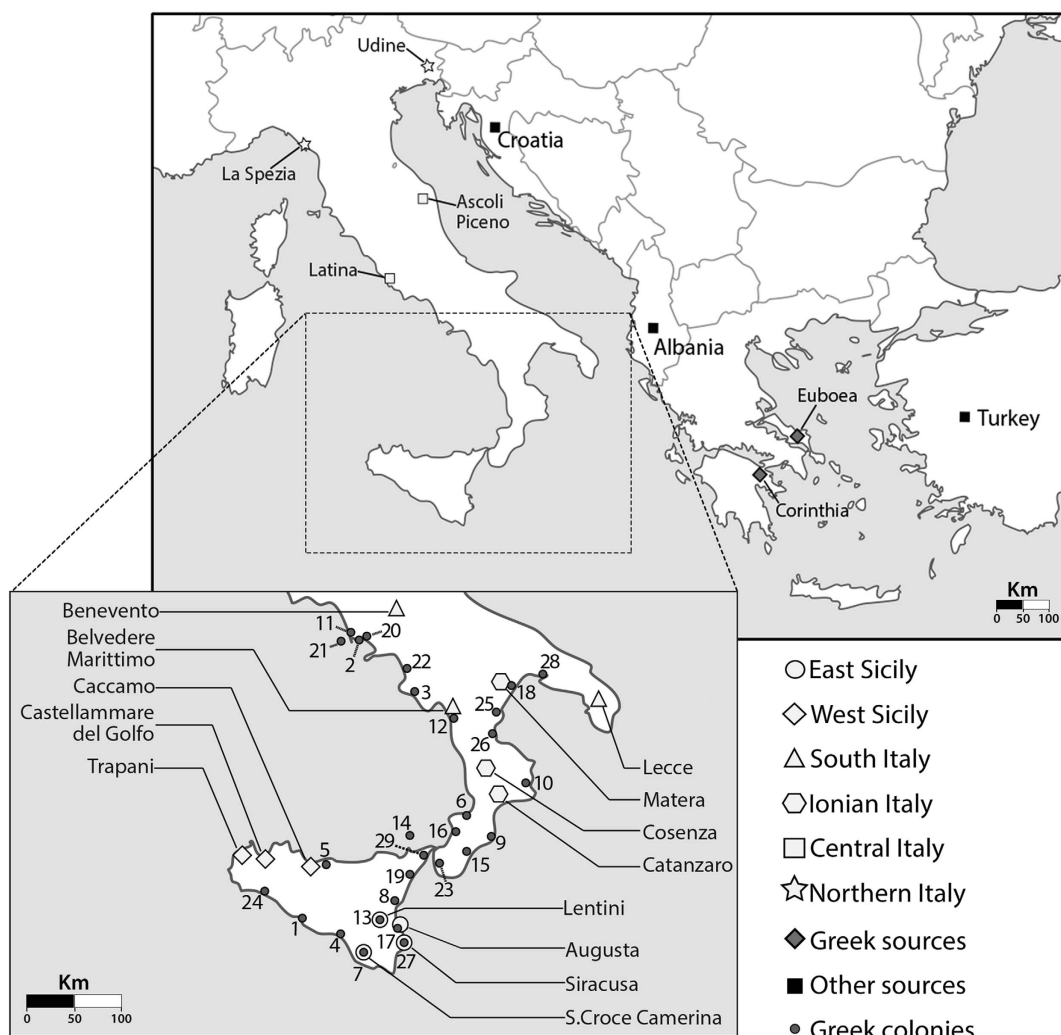


Figure 1 Map pointing the geographic origin of source, recipient and reference population samples. Primary and secondary colonies founded during the Archaic Age of the Hellenic expansion to South Italy and Sicily are also shown in grey dots.

Amplification kit (Applied Biosystems, Foster City, CA) and additional 9 YSTRs (DYS460, DYS388, YCA-II a/b, DYS461, DYS445, YGATA-A.10 and DYS413 a/b) by suitably designed multiplexed-PCR reactions (Supplementary Table S4). We finally assembled a haplotype data set based on 20 of the 26 analysed STR markers by excluding those STRs that in the PCR analysis co-amplify two loci and whose allele assignment to a defined locus was not possible (DYS385 a/b, YCA-II a/b and DYS413 a/b). A subset of samples ($N=304$) has been analysed for the hypervariable region I of mitochondrial DNA (mtDNA) using primers 15997L and 017H.

Data are available at <http://zenodo.org/>: DOI:10.5281/zenodo.15988 (Y markers); DOI:10.5281/zenodo.15987 (mtDNA markers).

Statistical analyses

An F_{ST} genetic distance matrix²⁷ was computed on Y haplogroup frequencies using the Arlequin package (version 3.5.1.2)²⁸ and graphically represented by a non-metric multidimensional scaling.²⁹ The analysis involved 18 population samples: the 6 Italian recipient groups, the 2 source samples from Greece (Euboea and Corinthia), the 3 reference samples (Turkey, Albania and Croatia) and 7 additional samples from Crete,^{30,31} mainland Greece³¹ and Lebanon.³² To make possible comparison across data sets genotyped with different single nucleotide polymorphism panels, haplogroups were pooled to the least basal common node on the ISOGG 2015 Y tree (<http://www.isogg.org/tree/>) for a total of 17 groups. Stress value suggests a non-random distribution of population samples in the bi-dimensional plot (0.0758, $P<0.05$).³³

Inferring pairs of Y haplotypes with GC ancestry

To identify pairs of Y haplotypes in the source and recipient populations with a time since the most recent common ancestor (TMRCA) compatible with the migrations from Greece to southern Italy and Sicily in the Archaic Period, we used Equation 31 (described by Walsh,²⁵ p. 907) as implemented in the software ASHES 1.1 (ashes.codeplex.com).³⁴ Briefly, TMRCA Bayesian posterior distributions were calculated for pairs of chromosomes separated by 0 to 2k mutational steps (where k is equal to the number of loci), assuming haplotypes composed by the set of 20 selected Y-STR loci (see above), a strict stepwise mutational model, a mutation rate of 3.09×10^{-3} per locus per generation (averaged values from Burgarella *et al*³⁵ and Ballantyne *et al*³⁶) and a lambda value of 0.0002 ($1/N$, where N =effective population size; here we used $N=5000$ in accordance with the study by Hammer³⁷). For each k , we explored the likelihood distribution at 102 (GC scenario; ~2750 years ago using 27 years per generation) and 300 (neolithic scenario; ~7500 years ago using 25 years per generation) generations (Supplementary Table S5). We corrected the male generation intervals calculated on present-day genealogies (31.9 years³⁸) according to the generalised reduction of life expectancy in pre-historic societies and the Y-based estimates calibrated in traslocated historical groups (25–30 years³⁹). To assess which interval of mutational differences between haplotypes is the most suitable to represent the GC contribute to Italy, we normalised each distribution and chose the mutational range within the likelihood inferred for the GC scenario, which minimised the overlapping (<0.4%) with the distribution inferred for the neolithic scenario. This range was identified as 8–12 mutational steps.

To estimate the 95% confidence boundaries, we approximated each distribution to a normal one, centred on the most likely TMRCA value. We then calculated the area below the posterior distribution comprised between the most likely value and the origin of the curve, and equated this value to 50% of the total likelihood curve. The 2.5–97.5% bounds of the distribution were identified as the TMRCA values comprising the 95% of the right and left hand sides likelihood of the curve.

Tracking genetic footprints of the GC legacy in Italy

To detect Greece-to-Italy genetic contributions from the Archaic Period (1000–400 BCE), we first performed pairwise comparisons at 20-locus YSTRs haplotypes between either sources (from Euboea, Corinth) or reference populations (from Croatia, Albania and Turkey) and putative recipient populations (20 Italian groups pooled in 6 geographical districts, Supplementary Table S1). Then, we calculated, for each population pair and within the same haplogroup, the proportion of haplotypes diverging for 8–12

mutational differences. To assess whether the enrichment in haplotype pairs matching this interval was significantly higher between sources and recipient populations than between sources and reference populations, we performed Fisher's exact tests with the Arlequin software v 3.5.²⁸ We applied the same procedure to both the whole sample (not filtered) and a subset of haplotypes (filtered), the latter obtained by removing all haplotypes from recipient populations with fewer than seven mutational differences from source population haplotypes, since the effect of more recent demography might overestimate this signal.

Computer simulations

To estimate the contribution of Greek colonisers to present-day southern Italian communities, we applied a hypothesis-testing approach based on the deviation from observed values of the haplotype divergence expected under different simulated scenarios. As measure of molecular divergence we used DHS, a distance measure that calculates the extent of exactly matching haplotypes between pairs of diverging pools.⁴⁰ The closer the divergence between pools of haplotypes, the lower the value of DHS, ranging from 0 (all haplotypes shared by the two populations) to 1 (no shared haplotypes).

As paternally inherited molecular markers we used a set of YSTRs selected on the number and the type which maximise the duration of DHS linearity with time. Six was the highest number that ensures DHS increases linearly within the last 300 generations (Supplementary Figure S1). Accordingly, we choose six loci: DYS393, DYS445, DYS456, DYS460, DYS461 and GATA-A10. The selected panel included loci characterised by tetrameric regular repeats, high and comparable estimated mutation rates (between 2.5 and 3.3×10^{-3} mutations per generation according to Burgarella *et al*³⁵ and Ballantyne *et al*³⁶) and no incomplete alleles. As maternally inherited molecular markers we used 360 bp hypervariable region I sequences, whose divergence, as measured by DHS, increases linearly over 300 generations (Supplementary Figure S2).

Different sets of parameters (see Supplementary Table S6) were modelled under a stochastic Markov chain Monte Carlo method as implemented in the software ASHES (<http://ashes.codeplex.com/>). As starting haplotype pools, we used source and recipient meta-populations obtained by reiterating the real data n times (Euboea and Corinth samples as source, a mix of southern Italian samples as recipient) until a final N_e equal to one-sixth of the current census size estimated according to the two demographic models was reached. For each simulation model, we considered two populations coming into contact at time t_0 and exchanging $M=N_e m$ haplotypes from the source to the recipient pool, where N_e is the effective size and m the fraction of migrants. From time t_1 , the two populations were allowed to evolve independently for 102 (Y haplotypes) or 110 (mt haplotypes) generations—that is, the time since the migration to Italy of the early colonisers from Greece (~2750 ya assuming, respectively, 27 and 25 years per generation). For each model, 100 iterations were performed and summary statistics of DHS values were calculated. We considered as varying parameters the initial effective size of source and recipient populations, the increment rate (0.00 ind/gen, stationary model; 0.01 individuals per generation, growth model), and the number of exchanged haplotypes M (0, 500, 1000 and 5000). Invariants were the mutation rate (0.0027 mut/site/gen for Y haplotypes; 0.0000041 mut/site/gen for mt haplotypes), haplotype diversity (0.90 and 0.80, respectively, for source and recipient Y haplotypes; 0.96 and 0.92, respectively, for source and recipient mt haplotypes) and DHS between source and recipient pools at t_0 (0.7 for Y haplotypes; 0.5 for mt haplotypes). The latter values are those expected between pool of haplotypes coming into contact 2750 ya after an initial divergence from an ancestral Anatolian pool some 6250 (150 gen) or 7500 (200 gen) ya under a model implying germ-line mutation rates as above and no size increment (data not shown).

The distributions of simulated DHS values were compared with empirical values calculated for each source/recipient pair of samples. Euboea and Corinth have been used as source samples, and Italians (pooled into six main geographic areas: West Sicily, East Sicily, Ionian Italy, South Italy, Central Italy and North Italy) have been used as recipient samples. The data were considered to fit the model when observed DHS values fell within 2 standard deviations (s.d.) by the mean of the simulated distribution.

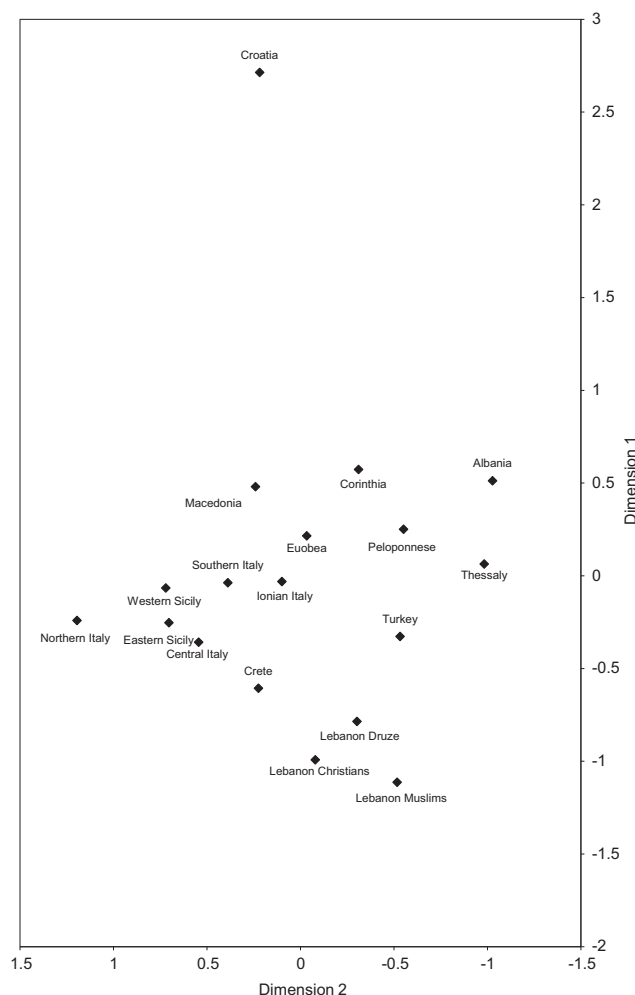


Figure 2 Non-metric multidimensional scaling bi-dimensional plot of F_{ST} pairwise genetic distances among investigated population samples and additional reference samples from the literature (Cretan and Lebanese) based on the frequencies of 17 Y haplogroups.

RESULTS

Population relationships

Y haplogroup frequencies are reported in Supplementary Table S7 while the overall pattern of inter-population genetic relationships is shown in Figure 2 and Supplementary Table S10. Cretan, mainland Greek and Lebanese samples were introduced to widen the spectrum of the historical players acting in south-eastern Mediterranean at the time of the GC, as proxies of non-Corinthian Dorian colonisers of South Sicily (since Crete contributed to the foundation of Gela and, in turn, Akragas), non-Euboean Ionian colonisers of South Italy and East Sicily, and Phoenician settlers in West Sicily, respectively. Looking at the reciprocal positions on the plot, little evidence of these historical events emerges, with the positioning reflecting geography rather than history. Accordingly, we observe higher genetic distance than that expected based on archaeological evidence between putative descendants of source (Greeks from Ionia, Corinthians and Cretan) and recipient (Sicilians) groups of the GC, as well as between the putative founders (Lebanese) of the Phoenician colonies in western Sicily (Motya, Panormos and Solus) and the present-day population.

Signatures of the Archaic Hellenic contribution

To detect genetic signatures of Greek migration in southern Italy and Sicily related compatible with the Archaic scenario, we compared fractions of haplotype pairs within the 8–12 mutational range, or 'GChp', with the same fractions obtained by using Albanian, Croatian and Turkish samples as reference sources (Table 1a).

Samples from East Sicily, West Sicily, South and Central Italy showed significant ($P < 0.01$) enrichment of GChps when the Greek sample from the Euboea Island was compared with Corinthia and reference sources. Except in West Sicily and Central Italy, this enrichment remained highly significant even after correcting for multiple tests ($P < 0.05$).⁴¹ Conversely, when considering Corinthia against other reference sources, none of the recipient samples showed a full set of significant values. The comparisons involving recipients *versus* Albania most commonly showed a lack of significant enrichment in GChps with respect to Euboea and Corinthia. We reasoned that contacts either between sources and recipients or between sources after the GC, that is, during the Classical and Christian periods, might have contributed to increase the GChps rate. Thus, to provide more stringent conditions for haplotype identification, we excluded all haplotypes with a molecular distance less than seven mutational steps. This 'filtered' data set confirmed the pattern observed with the less stringent criteria for the East Sicily/Euboea pair, which showed significant enrichment in GChps in two out of three comparisons even after the Bonferroni correction (Table 1b). The results for West Sicily and South Italy did not hold statistical significance when a Bonferroni correction was applied. None of the other Italian recipients showed a full set of significant enrichment with Euboea or Corinthia.

Estimating Greek contribution

The number of GChps identified using the suggested molecular distance cannot be used either to directly estimate the current Greek legacy in Italian populations or to provide an indication of the original demographic contribution. Nevertheless, this approach helped us to identify populations (East Sicily, and, to a lesser degree, West Sicily and South Italy), that are characterized by a significant association with Greek populations derived during the time window of interest.

To quantify the original demographic impact of the Greek settlers inferable from present-day Y chromosome variability, we explored two main census scenarios using a simulation-based approach. In the first scenario, high count, we based our model on the demographic estimates of Beloch,^{18,19} who suggested a census size of 1.35 million people for Sicily and of 3 million for Greece at the time of the Hellenic colonisation in the Archaic Period. In the second scenario, low count, we modelled population size estimates that were smaller by an order of magnitude.²¹ Nevertheless, the two scenarios have similar source/recipient effective size ratios (S/R). If we assume that the proportions of past Sicilian, Euboean and Peloponnese census and male (and female) effective population sizes is one-sixth of the current census size, we estimate a S/R of 3.65 for the high-count model and a S/R of 3.75 for the low-count model. Simulation results are reported on Figure 3 and Supplementary Table S8. When considering Y-STR haplotypes, the observed DHS value between Euboea and East Sicily (0.5353) is compatible with an effective number of migrants ranging between 500 and 5000, clearly rejecting larger contributions (10 000), irrespective of the scenario considered. The DHS value obtained for hypervariable region I haplotypes (0.5995) supports an effective number of migrants between 500 and 1000, with larger contributions clearly excluded. When the NRY- and mtDNA-based estimates are paired according to the demographic model, the male-to-female

Table 1a Significance levels of haplotype enrichment within 8–12 mutational steps

	<i>Euboea vs</i>			<i>Corinthia vs</i>			<i>Euboea vs Corinthia</i>
	<i>Croatia</i>	<i>Albania</i>	<i>Turkey</i>	<i>Croatia</i>	<i>Albania</i>	<i>Turkey</i>	
East Sicily	0.00000	0.00002	0.00000	0.02497	0.85592	0.00006	0.00013
West Sicily	0.00001	0.01098	0.01487	0.01734	0.60043	0.72580	0.05869
South Italy	0.00000	0.00186	0.00257	0.01804	0.51419	0.60002	0.01727
Ionian Italy	0.00000	0.30148	0.00230	0.00005	0.89556	0.09685	0.19161
Central Italy	0.00000	0.04539	0.00914	0.00000	0.54767	0.21634	0.22274
North Italy	0.00495	0.23218	0.00283	0.20377	1.00000	0.13417	0.15605
After Bonferroni correction:							
East Sicily	0.00000	0.00008	0.00000	0.09988	1.00000	0.00024	0.00052
West Sicily	0.00004	0.04392	0.05948	0.06936	1.00000	1.00000	0.23476
South Italy	0.00000	0.00744	0.01028	0.07216	1.00000	1.00000	0.06908
Ionian Italy	0.00000	1.00000	0.00920	0.00020	1.00000	0.38740	0.76644
Central Italy	0.00000	0.18156	0.03656	0.00000	1.00000	0.86536	0.89096
North Italy	0.01980	0.92872	0.01132	0.81508	1.00000	0.53668	0.62420

Values with $P < 0.05$ are in bold.**Table 1b** Significance levels of haplotype enrichment within 8–12 mutational steps after removing haplotypes showing matches within the 0–7 mutations range

	<i>Euboea vs</i>			<i>Corinthia vs</i>			<i>Euboea vs Corinthia</i>
	<i>Croatia</i>	<i>Albania</i>	<i>Turkey</i>	<i>Croatia</i>	<i>Albania</i>	<i>Turkey</i>	
East Sicily	0.03302	0.00624	0.00011	0.65675	1.00000	0.16146	0.00520
West Sicily	0.04113	0.01972	0.04546	0.78290	0.58171	0.78340	0.13875
South Italy	0.00027	0.17368	0.07584	0.00009	0.06603	0.02373	0.85785
Ionian Italy	0.00030	0.82917	0.11442	0.08607	0.25120	1.00000	0.10902
Central Italy	0.00358	0.64544	1.00000	0.00767	0.81960	1.00000	1.00000
North Italy	0.01634	0.16904	0.42036	0.30721	1.00000	1.00000	0.27607
After Bonferroni correction:							
East Sicily	0.13208	0.02496	0.00044	1.00000	1.00000	0.64584	0.02080
West Sicily	0.16452	0.07888	0.18184	1.00000	1.00000	1.00000	0.55500
South Italy	0.00108	0.69472	0.30336	0.00036	0.26412	0.09492	1.00000
Ionian Italy	0.00120	1.00000	0.45768	0.34428	1.00000	1.00000	0.43608
Central Italy	0.01432	1.00000	1.00000	0.03068	1.00000	1.00000	1.00000
North Italy	0.06536	0.67616	1.00000	1.00000	1.00000	1.00000	1.00000

Values with $P < 0.05$ are in bold.

migrant ratio ranged between 1:1 and 2:1 under a population growth model and between 2:1 and 10:1 under a constant population size model.

Lineage-based demographic estimates

Previous investigations^{14,15} have suggested that the Y chromosome lineage E-V13 is a marker of the Hellenic contribution in the Mediterranean. To test the validity of this approach, we repeated the enrichment test described above by considering only haplotypes belonging to the E-V13 lineage. Given the relatively low frequency of this haplogroup in areas outside the Balkan peninsula, only the East Sicily sample provided a size of NRY haplotypes ($N = 20$) large enough to perform meaningful comparisons. All the 20 E-V13 haplotypes in the sample from East Sicily had matches in the 8–12 mutational range when compared with GC and reference sources. Accordingly, no enrichment in GChps was found except *versus* Turkey (Fisher's exact test, $P < 0.01$). When haplotype pairs with mismatches of 0–7 steps

were removed (F data set), both Croatian and Turkish samples showed an increased relative number of GChps (respectively, 9 and 7) with respect to the other source samples (2 in Albanian and Corinthian samples, 1 in the Euboean sample).

We further explored the impact of haplotypes belonging to specific lineages by calculating the fraction of GChps belonging to the various single nucleotide polymorphism-defined lineages contributing to the overall enrichment (Supplementary Table S9). E-V13 is the major contributor for all the sources excluding Croatia. The contribution of E-V13 in Euboea and Corinthia was much lower than in the three reference sources (range 12.6–17.1 *versus* 17.5–22.8%). Similarly, we evaluated the contribution of E-V13 GChps in the F data set (Supplementary Table S9). As such, E-V13 haplotypes are no longer the major fraction of GChps and were under-represented in Corinthia (3.4%) and Euboea (1.1%). Figure 4 clearly shows that for these two candidate GC sources the largest quote of E-V13 haplotypes pairs did not fall in the 8–12 but in the 1–4 mutational step interval. Moreover,

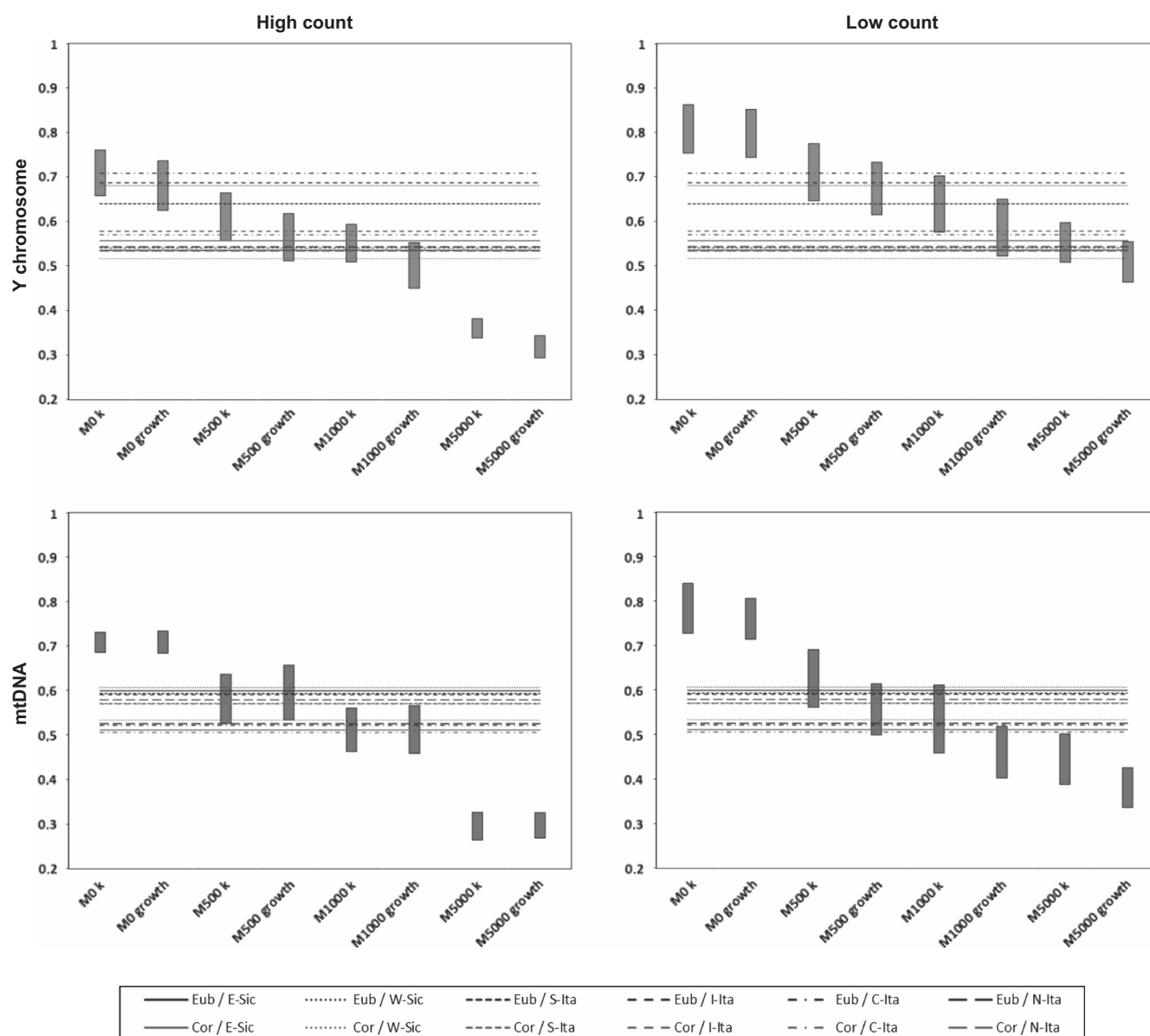


Figure 3 Plots of observed (lines) and expected (bars) DHS values for each simulated evolutionary scenario. MO, M500, M1000, M5000=Number of migrants; k =stationary model, growth=expansion model with increment $\lambda=0.01$ ind/gen. High count=census sizes as in the study by Beloch¹⁸; low count=census sizes as in the study by Turchin²¹. Dark gray lines/bars=results involving the Euboean sample as source population; light gray lines/bars=results involving the Corinthian sample as source population. A certain parameter set was considered as fitting the model when the observed DHS fell within a two standard deviations interval (bars) of the simulated distribution at time $t=102$ (Y haplotypes) and $t=110$ (mtDNA haplotypes) generations.

it demonstrates that only haplotype pairs within the Albania sample reached the highest peak within the 8–12 range.

DISCUSSION

Evidence of Hellenic genetic echoes in Italy

The history of the European continent has been characterised by a large number of migration and admixture events.⁴² The peopling of the Mediterranean is a clear example of this complexity. Hence, not here the use of descriptive approaches aiming at summarising the observed genetic variation can easily miss signatures related to a given event.

Aware of these limitations, we attempted to recover genetic signals related to the Greek colonisation in southern Italy by analysing samples specifically collected to provide information on the source

and recipient populations actually involved. None of the Italian populations showed a closer affinity with Greek and Greek-related sources when Y chromosome data was analysed using multidimensional scaling analyses (Figure 2). There are many explanations for this: limited historical migration between the two countries; lack of continuity between original and present-day source populations with current samples not being a good proxy of the ancestral populations they come from; the signal of the ancestry could be confounded by more recent or more ancient events. Nevertheless, when we used an approach designed to take into account the mutational process, we recovered a signature of the Greek Contribution to Sicily during the Archaic Period. A first-level analysis based on the Bayesian posterior distribution of mutational steps compatible with the former colonisation phase (GChps) showed that the most evident signal was in East

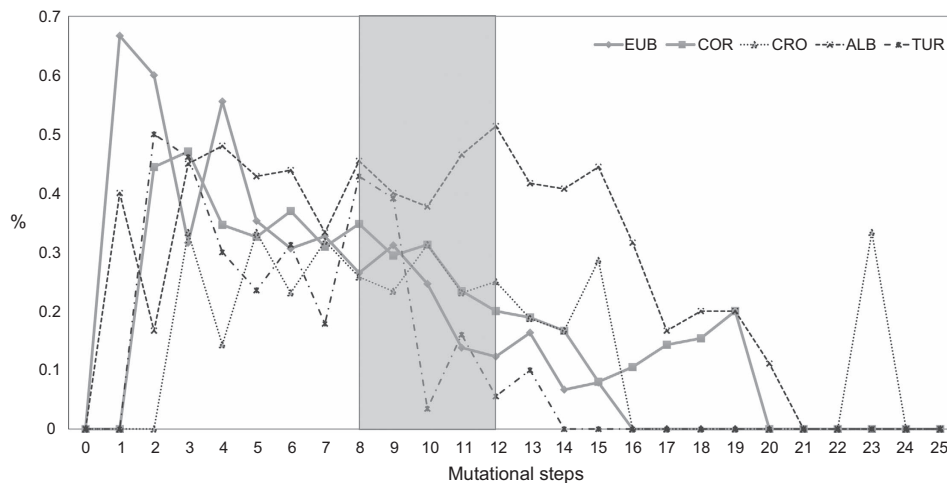


Figure 4 Relative percentage of mutational steps between E-V13 haplotypes pairs in source and reference population samples.

Sicily, but this was also found to have parallels, or have diffused into neighbouring regions of West Sicily and South Italy. A second-level analysis, performed under more stringent conditions, again detected a signal in East Sicily. The lack of similar signals in other areas known to have been colonised by Greek migrants (eg, Ionic Italy) can be explained by inadequate source samples (ie, the Achaia region is not represented in our data set), as well as by either a lower demographic impact or subsequent population discontinuity. We note that many of the pairwise comparisons were not significant due to high background signals from the Albanian sample, this being particularly noticeable when the Corinthian sample was involved. The easiest explanation for this is the close genetic affinity observed among the two areas (see Figure 2) deriving from direct and/or indirect gene flow. For example, it is known that Greek colonies in present Albanian territory, such as Apollonia and Epidamnos, were founded by Corinthians⁴ and that the area around Korinth was settled by southern Albanian orthodox Christians between the thirteenth and sixteenth century, the descendants of whom are identified as Arvanites.⁴³ During the sample collection, attention was given to this issue: individuals who self-reported as Arvanite were excluded from the analysis. Nevertheless, some Arvanite ancestry will have been unreported and may have affected the results.

Despite the multiple alternative explanations for historical gene flow, it is relevant to stress here that a signature specifically related to the Euboea island in East Sicily was consistently found at different levels of analysis, in line with the historical and archaeological evidences,^{2–4} attesting an extended and numerically important Greek presence in this region.

Sex-biased gene flow

The numerical dimensions of the migration from Greece that resulted in the establishment of Hellenic colonies have been debated by scholars for centuries (see the study by Scheidel⁸ among others).

The signal from East Sicily points towards the lower end of the size spectrum proposed by historical demographers, with values in the order of thousands breeding men and few hundreds breeding women. From this perspective, our results are then compatible with the hypothesis that the migration and settlement process was driven by males. Interestingly this is one of the few cases of sex-biased gene flow skewed towards an increased male instead of female contribution.⁴⁴

It is obvious that such numbers refer to the colonists who arrived in East Sicily as inferred from the descendants still living today. It is also

worth stressing here that such estimates should not be taken as absolute but considered as indicative of the scale of the contribution. A possible flaw in these results could be also due to the sex-differential migratory rate after the first settlement. It is known that patrilocality is a commonplace in continental Italy and Sicily. This may have facilitated the diffusion of mtDNA variation at a larger scale than Y chromosome variation, thus lowering the probability of finding local female genetic signatures of the GC (the study by Heyer⁴⁴ but see the study by Marks⁴⁵).

Lineage-based estimations

The genetic contribution of a given source within a defined historical scenario has often been estimated using the number of chromosomes assigned to given haplogroups assumed to have a specific geographic/ethnic origin. In relation to the GC of the Mediterranean, the lineage defined by the E-V13 marker has been used to estimate the Hellenic contribution to the Sicilian gene pool.¹⁴ By assuming that all E-V13 chromosomes have a Hellenic origin, authors estimated a contribution of ~37% to the population in Sicily. The reconstruction of a STR-based network linking the Sicilian modal haplotype and its one-step neighbouring haplotypes provided a TMRCA of about 2380 years before present, with a 95% confidence ranging between 675 and 6940 years ago. More recent contributions and differential origins are expected to affect such estimates, but attempts to mitigate the impact of these phenomena were not implemented.

In this study we highlighted that, when alternative sources were taken in consideration, E-V13 did not show any specificity as a marker of the Hellenic contribution. The signal that we found using the full set of haplotypes within a limited range of mutational distance disappeared when only E-V13 unbounded haplotypes are considered, becoming even significant for other non-Greek sources when a filter for recent gene flow was applied. The contribution of E-V13 from Corinthia and Euboea, in fact, reaches its peak well before the 8–12 mutational steps range (Figure 4), while in this range and samples it showed minor relative frequency.

These findings suggest a poor association between the Y haplogroup E-V13 and the East-to-West GC migratory waves. The effects of more recent gene flow or sampling bias, may have masked the original E-V13 signal from Greece. At any rate, our results caution against the use of specific lineage-based approaches to test for hypothesised population contributions and underline the need for a more targeted approach to explain the occurrence of given haplotypes within

a population, providing tests of alternative hypotheses, a wide spectrum of reference samples and mutation-limited inference methodology.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This project was funded by the British Academy (BARDA-47870; CC). The authors would like to acknowledge all the participants who offered their biological samples for analyses. We thank Giacomo De Leo and Anna Flugy (the Università di Palermo) and the staff of the Servizio di Medicina Trasfusionale dell'Ospedale Muscatello di Augusta, AVIS Comunale di Santa Croce camerina, UOC Medicina Trasfusionale, Ospedale di Lentini, Unità Operativa Complessa di Medicina Trasfusionale, Azienda Ospedaliera Umberto I and Siracusa for assistance in sample collection. ST is grateful to Domenico Accorinti and Domitilla Campanile for their suggestions. The Croatian collection was funded by grants from the Medical Research Council (UK), European Commission Framework 6 project EUROSPAN (Contract No LSHG-CT-2006-018947) and Republic of Croatia Ministry of Science, Education and Sports research grants to IR (108-1080315-0302).

- Graham AJ: The colonial expansion of Greece; in Boardman J, Hammond NG (eds): *The Cambridge Ancient History: The Expansion of the Greek World, Eighth to Sixth Centuries*. Cambridge University Press: New York, 1982, pp 83–162.
- Pugliese Carratelli G: An outline of the political history of the Greeks in the West; in Pugliese Carratelli G (ed): *The Western Greeks*. Bompiani: Milano, Italy, 1996, pp 141–176.
- Guzzo PG: *Fondazioni Greche: L'Italia meridionale e la Sicilia (VIII e VII sec. a.C.)*. Roma: Carocci Editore. 2011.
- Boardman J: *The Greeks Overseas*. Thames & Hudson: London, 1980.
- Yntema D: Archaeology and the *Origo* Myths of the Greek *Apoikiai*. *Ancient West East* 2011; **10**: 243–266.
- Dunbabin TJ: *The Western Greeks*. Thames & Hudson: Oxford, 1948.
- Faure P: *La vie Quotidienne des Colons Grecs*. Hachette: Paris, 1978.
- Scheidel W: The Greek demographic expansion: models and comparisons. *J Hell Stud* 2003; **123**: 120–140.
- Piazza A, Cappello N, Olivetti E *et al*: A genetic history of Italy. *Ann Hum Genet* 1988; **52**: 203–213.
- Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton University press: Princeton, 1994.
- Capelli C, Brisighelli F, Scarnicci F *et al*: Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter. *Mol Phylogenet Evol* 2007; **44**: 228–239.
- Brisighelli F, Álvarez-Iglesias V, Fondevila M *et al*: Uniparental markers of contemporary Italian population reveals details on its pre-Roman heritage. *PLoS One* 2012; **7**: e50794.
- Boattini A, Martinez-Cruz B, Sarno S *et al*: Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. *PLoS One* 2013; **29**: e65441.
- Di Gaetano C, Cerutti N, Crobu F *et al*: Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome. *Eur J Hum Genet* 2009; **17**: 91–99.
- King RJ, Di Cristofaro J, Kouvatsi A *et al*: The coming of the Greeks to Provence and Corsica: Y-chromosome models of archaic Greek colonization of the western Mediterranean. *BMC Evol Biol* 2011; **11**: 69.
- Edmonds CA, Lillie AS, Cavalli-Sforza LL: Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci USA* 2004; **101**: 975–979.
- Klopfstein S, Currat M, Excoffier L 2006 The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* **23**: 482–490.
- Beloch KJ: *Die Bevölkerung der Griechisch-Römischen Welt*. Duncker: Leipzig, Germany, 1886.
- McEvedy C, Jones R: *Atlas of World Population History*. Penguin: Middlesex, England, 1978.
- Hansen MH: *The Shotgun Method: The Demography of the Ancient Greek City-State Culture*. University of Missouri Press: Columbia-Missouri, USA, 2006.
- Turchin P, Scheidel W: Coin hoards speak of population declines in ancient Rome. *Proc Natl Acad Sci* 2009; **106**: 17276–17279.
- Ferri G, Tofanelli S, Alù M *et al*: Y-STR variation in Albanian populations: implications on the match probabilities and the genetic legacy of the minority claiming an Egyptian descent. *Int J Legal Med* 2010; **124**: 363–370.
- Rudan I, Marusić A, Janković S *et al*: "10001 Dalmatians": Croatia launches its national biobank. *Croat Med J* 2009; **50**: 4–6.
- Richards M, Macaulay V, Hickey E *et al*: Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 2000; **67**: 1251–1276.
- Walsh B: Estimating the time to the MRCA for the Y chromosome or mtDNA for a pair of individuals. *Genetics* 2001; **158**: 897–912.
- The Y Chromosome Consortium: A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002; **12**: 339–348.
- Slatkin M: A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 1995; **139**: 457–462.
- Excoffier L, Lischer HE: Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010; **10**: 564–567.
- Kruskal J: Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 1964; **29**: 28–42.
- Martinez L, Underhill PA, Zhivotovsky LA *et al*: Paleolithic Y-haplogroup heritage predominates in a Cretan highland plateau. *Eur J Hum Genet* 2007; **15**: 485–493.
- King RJ, Ozcan SS, Carter T *et al*: Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic. *Ann Hum Genet* 2008; **72**: 205–214.
- Zalloua PA, Xue Y, Khalife J *et al*: Y-chromosomal diversity in Lebanon is structured by recent historical events. *Am J Hum Genet* 2008; **82**: 873–882.
- Sturrock K, Rocha J: A multidimensional scaling stress evaluation table. *Field Methods* 2000; **12**: 49–60.
- Capelli C, Onofri V, Brisighelli F *et al*: Moors and Saracens in Europe: estimating the medieval North African male legacy in southern Europe. *Eur J Hum Genet* 2009; **17**: 848–852.
- Burgarella C, Navascués M: Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. *Eur J Hum Genet* 2011; **19**: 70–75.
- Ballantyne KN, Goedbloed M, Fang R *et al*: Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic Implications. *Am J Hum Genet* 2010; **87**: 341–353.
- Hammer MF: A recent common ancestry for human Y chromosomes. *Nature* 1995; **378**: 376–378.
- Helgason A, Hrafnkelsson B, Gulcher JR *et al*: A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *Am J Hum Genet* 2003; **72**: 1370–1388.
- Bonné-Tamir B, Korostishevsky M, Redd AJ *et al*: Maternal and paternal lineages of the Samaritan isolate: mutation rates and time to most recent common male ancestor. *Ann Hum Genet* 2003; **67**: 153–164.
- Tofanelli S, Bertocini S, Castri L *et al*: On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. *Mol Biol Evol* 2009; **26**: 2109–2124.
- Bonferroni CE: Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 1936; **8**: 1–62.
- Hellenthal G, Busby GB, Band G *et al*: A genetic atlas of human admixture history. *Science* 2014; **343**: 747–751.
- Hall JM: *Ethnic Identity in Greek Antiquity*. Cambridge University Press: Cambridge, 2000.
- Heyer E, Chaix R, Pavard S *et al*: Sex-specific demographic behaviours that shape human genomic variation. *Mol Ecol* 2012; **21**: 597–612.
- Marks SJ, Levy H, Martinez-Cadenas C *et al*: Migration distance rather than migration rate explains genetic diversity in human patrilocal groups. *Mol Ecol* 2012; **21**: 4958–4969.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)